

# Geophysical Research Letters

## RESEARCH LETTER

10.1029/2020GL089102

### Key Points:

- Tropical cyclone intensity change is approximately linearly correlated with surplus inner-core precipitation above a threshold value
- The threshold value can be determined from neutral TCs of nearly constant intensity, and it increases linearly with TC intensity
- Machine learning with combined environmental factors and surplus precipitation significantly improves TC rapid intensification forecasts

### Supporting Information:

- Supporting Information S1

### Correspondence to:

H. Su,  
hui.su@jpl.nasa.gov

### Citation:

Su, H., Wu, L., Jiang, J. H., Pai, R., Liu, A., Zhai, A. J., et al. (2020). Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophysical Research Letters*, 47, e2020GL089102. <https://doi.org/10.1029/2020GL089102>

Received 8 JUN 2020  
Accepted 6 AUG 2020

## Applying Satellite Observations of Tropical Cyclone Internal Structures to Rapid Intensification Forecast With Machine Learning

Hui Su<sup>1</sup> , Longtao Wu<sup>1</sup> , Jonathan H. Jiang<sup>1</sup> , Raksha Pai<sup>2</sup> , Alex Liu<sup>3</sup>, Albert J. Zhai<sup>4</sup> , Peyman Tavallali<sup>1</sup> , and Mark DeMaria<sup>5</sup>

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, <sup>2</sup>IBM Global Business Services, Armonk, NY, USA, <sup>3</sup>RMDS Lab, Pasadena, CA, USA, <sup>4</sup>Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA, <sup>5</sup>National Hurricane Center, Miami, FL, USA

**Abstract** Tropical cyclone (TC) intensity change is controlled by both environmental conditions and internal storm processes. We show that TC 24-hr subsequent intensity change (DV24) is linearly correlated with the departures in satellite observations of inner-core precipitation, ice water content, and outflow temperature from respective threshold values corresponding to neutral TCs of nearly constant intensity. The threshold values vary linearly with TC intensity. Using machine learning with the inner-core precipitation and the predictors currently employed at the National Hurricane Center (NHC) for probabilistic rapid intensification (RI) forecast guidance, our model outperforms the NHC operational RI consensus in terms of the Peirce Skill Score for RI in the Atlantic basin during 2009–2014 by 37%, 12%, and 138% for DV24  $\geq$  25, 30, and 35 kt, respectively. Our probability of detection is 40%, 60%, and 200% higher than the operational RI consensus, while the false alarm ratio is only 4%, 7%, and 6% higher.

**Plain Language Summary** Improving the predictive skill of TC intensity change, especially rapid intensification (RI), has been a top priority of the National Hurricane Center (NHC). We show TC 24-hr subsequent intensity change (DV24) is approximately linearly correlated with the departures in satellite observations of inner-core precipitation, ice water content, and outflow temperature from respective threshold values corresponding to neutral TCs of nearly constant intensity. The threshold values vary linearly with TC intensity. Using machine learning techniques, we construct a statistical model for RI forecast with a combination of the predictors included in the NHC probabilistic forecast guidance and satellite observations of storm internal structure. The overall predictive skill of our model for 2009–2014 RI in the Atlantic basin is better than the NHC operational RI consensus by 37%, 12%, and 138% for DV24  $\geq$  25, 30, and 35 kt, respectively. The probability of detection for RI by our model is 40%, 60%, and 200% higher than the operational RI consensus, while the false alarm ratio is only 4%, 7%, and 6% higher.

## 1. Introduction

Improving tropical cyclone (TC) intensity forecast skill has tremendous socioeconomic value. However, progress in reducing the TC intensity forecast error has been moderate compared to improvements in TC track forecasts (DeMaria et al., 2014). Rapid intensification (RI), defined as 24-hr TC maximum sustained surface wind speed increase greater than 30 knots (kt, 1 kt = 0.514 m s<sup>-1</sup>), is especially difficult to predict and improving RI forecast skill has been the highest priority of the National Hurricane Center (NHC) (Rappaport et al., 2012).

It is well known that TC intensity change is sensitive to environmental conditions. The Statistical Hurricane Intensity Prediction Scheme (SHIPS) Rapid Intensification Index (RII) employed operationally at the NHC was developed primarily using the environmental parameters that exhibit statistically significant differences at the 99.9% level between the RI and non-RI cases, such as sea surface temperature (SST), lower and middle troposphere moisture, vertical wind shear, and upper tropospheric divergence (Kaplan & DeMaria, 2003; Kaplan et al., 2010, hereafter K10; Kaplan et al., 2015, hereafter K15). On the other hand, internal storm dynamics is also recognized as an important factor in RI processes, given favorable environmental conditions (Hendricks et al., 2010). Some studies suggest that intense convective bursts or hot towers are crucial to RI (e.g., Chen et al., 2018; Chen & Zhang, 2013; Hazelton et al., 2017; Molinari & Vollaro, 2010; Rogers

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

et al., 2015; Stevenson et al., 2014), while other studies argue that symmetric convective heating is more conducive to RI (Jiang, 2012; Nolan et al., 2007; Nolan & Grasso, 2003; Shapiro & Willoughby, 1982; Willoughby, 1990). In the SHIPS RII model, storm internal processes such as the strength and symmetry of convection are crudely represented by the coldness and standard deviation of geostationary infrared brightness temperature (BT) within 50- to 200-km radius (K15).

Numerous studies show that satellite measurements of inner-core precipitation rate and convective structures estimated from radar reflectivity or microwave BT are correlated with TC intensity and intensification rate (e.g., Cecil & Zipser, 1999; Fischer et al., 2018; Jiang, 2012; Jiang et al., 2019; Jiang & Ramirez, 2013; Kieper & Jiang, 2012; Rozoff et al., 2015; Shimada et al., 2018; Zagrodnik & Jiang, 2014), indicating the potential value of these observations in operational forecasts. Fischer et al. (2018) introduced an anomalous BT relative to the mean BT averaged for all TCs of the same intensity in order to decouple the dependence of BT on TC current intensity and future intensity change. They showed that the normalized BT could be a skillful predictor for RI (Fischer et al., 2018). However, an effective framework for transitioning these satellite observations into operational forecasts has not been developed.

Moreover, satellite retrievals of cloud, water vapor, and temperature profiles reveal three-dimensional storm structures and their environments. Wu et al. (2012) analyzed the relationship of environmental relative humidity (RH) with TC intensity and intensification rate using the RH observations from Atmospheric Infrared Sounder (AIRS) on Aqua. Wu and Soden (2017) examined the ice and liquid water content (I/LWC) from CloudSat and found strengthening TCs generally have greater IWC but not necessarily greater LWC than weakening TCs. Based on the TC potential intensity theory, outflow temperature in the tropical tropopause layer (TTL) is an important factor in determining TC maximum intensity (Emanuel et al., 2013; Wang et al., 2014). Hence, the TTL temperature measured by Aura Microwave Limb Sounder (MLS) may be related to TC intensity change. Due to their infrequent synoptic sampling frequency, measurements from polar-orbiting satellites are not widely used in operational statistical TC intensity guidance models. With recent progress in miniaturized satellites for remote sensing, it is conceivable that more frequent observations of 3-D storm structures will be available in the future (e.g., Blackwell et al., 2013). Hence, it is necessary to examine whether satellite retrievals of storm structures are useful for objective TC intensity forecast guidance.

This study aims to demonstrate the predictive capability of satellite observations of storm internal structures for TC forecasting by incorporating them in a statistical RI forecast model with machine learning (ML) techniques. It also evaluates the utility of the ML techniques compared with NHC's current objective RI guidance. We first present empirical relationships between these measurements and future TC intensity change and establish the observational basis for their use in operational forecast model. Then we show the predictive skill of the ML model for RI episodes in the Atlantic and Eastern North Pacific (EPAC) basins with combined SHIPS RII predictors and new predictors derived from the satellite observations. The ML model performance is evaluated against the NHC operational RI forecast guidance for 2009–2014 TCs.

## 2. Data and Methods

We examine satellite observations of surface precipitation rate, IWC, and the TTL temperature in this study. We sort these variables by TC current intensity ( $V$ ) and subsequent 24-hr intensity change (DV24). Based on TC maximum sustained surface wind speed from the Best Track archive from the NHC and the Joint Typhoon Warning Center (JTWC), four intensity groups are defined including tropical depression (TD,  $V \leq 33$  kt), tropical storm (TS,  $33 \text{ kt} < V \leq 63$  kt), Category 1–2 hurricanes (H12,  $63 \text{ kt} < V < 96$  kt), and Category 3–5 hurricanes (H345,  $V \geq 96$  kt). The Best Track archive reports TC intensity every 6 hr with 5 kt resolution. Four intensification rate groups are defined according to the value of DV24: weakening (W) ( $\text{DV24} < -5$  kt), neutral (N) ( $-5 \text{ kt} \leq \text{DV24} \leq 5$  kt), slowly intensifying (SI) ( $5 \text{ kt} < \text{DV24} < 30$  kt), and rapidly intensifying (RI) ( $\text{DV24} \geq 30$  kt). Only TCs in the Northern Hemisphere are considered in this study.

Surface precipitation is taken from the Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation product (Huffman et al., 2007). It has a 3-hr temporal resolution and  $0.25^\circ \times 0.25^\circ$  spatial resolution covering  $50^\circ\text{S}$  to  $50^\circ\text{N}$  from 1998 to 2014, which contains 26,879 samples (Table S1 in the supporting information). TRMM precipitation within 3 hr of the Best Track synoptic times are averaged. The CloudSat I/LWC data

for TCs were archived in the CloudSat TC (CSTC) database (Tourville et al., 2015). The Level 2 I/LWC has a resolution of 1.7 km along track and 1.3 km cross track. We use the overpasses with the point of closest approach (PCA) within 50 km of the TC center from 2006 to 2016, resulting in 294 samples (Table S2). The MLS temperature product has a resolution of  $\sim 165$  km along track and  $\sim 12$  km cross track (Schwartz et al., 2008) and only the overpasses with the PCA within 150 km of storm center are used, resulting in 467 samples from 2004–2017 (Table S3). The Best Track TC intensity and DV24 are interpolated onto the CloudSat and MLS measurement times.

As CloudSat and MLS data contain limited TC samples, we also analyze the ice water path (IWP) and TTL temperature from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis to increase sample size and statistical significance of the composite results. The MERRA-2 3-hourly reanalysis has a resolution of  $0.625^\circ \times 0.5^\circ$  (Gelaro et al., 2017). There are 44,728 TC samples from 1998 to 2017 over global oceans coincident with the Best Track TC data (Table S1).

The data sets and methods for the ML modeling are described in section 3.2.

### 3. Results

#### 3.1. Basis for Satellite Observations of Storm Internal Structures as RI Predictors

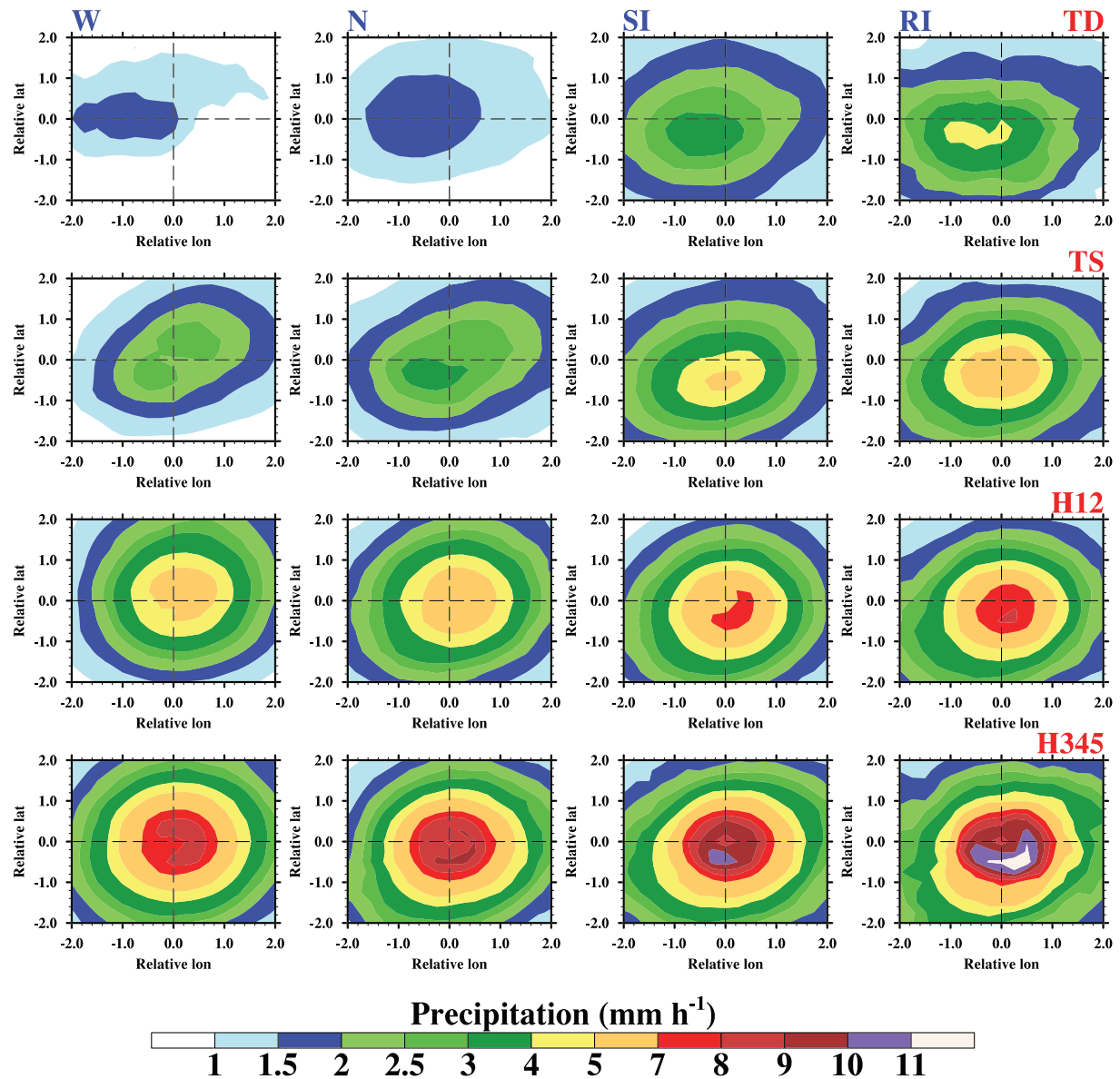
Figure 1 depicts the horizontal distribution of composite surface precipitation in storm-centered coordinates for different TC intensity and intensification rate groups. In each intensification rate group, the inner-core precipitation rate generally increases with TC intensity. Similarly, in each intensity group, the inner-core precipitation rate increases with intensification rate. In this study, the inner-core loosely refers to the area within about 100-km radius of storm center and our results are not very sensitive to the exact distance to the storm center constrained by the resolution of TRMM precipitation product. There appears to be some azimuthal asymmetry in the composite precipitation distribution, especially for major hurricanes (H345), although it is unclear how significantly the azimuthal asymmetry affects TC intensity change. A wavenumber analysis of the storm-centered precipitation field might be helpful (Bhalachandran et al., 2019), although, for simplicity, that is not considered in this study.

Displaying the mean precipitation rate within 100 km of storm center ( $P$ ) as a function of TC current intensity ( $V$ ) for the four intensification rate (DV24) groups (Figure 2a), we observe four nearly parallel semilinear curves that shift upward for increasing DV24. Defining a surplus  $S \equiv P - P_N$ , where  $P_N$  is the inner-core precipitation rate for neutral TCs of the same intensity and increases approximately linearly with  $V$ , DV24 would be an increasing function of  $S$  for all values of  $V$ . Figure 2b displays DV24 at 5-kt intervals against the composite mean  $S$  for all TCs and for three TC intensity groups. It is striking that all four curves in Figure 2b collapse to one with the curve for hurricanes shifting slightly to the left. This suggests that the composite relationship between DV24 and  $S$  is universal for all TCs of different intensities. For intensifying storms ( $DV24 > 5$  kt), DV24 increases approximately linearly with  $S$ . For weakening storms ( $DV24 < -5$  kt), DV24 is insensitive to  $S$ , suggesting other processes including asymmetric convective heating may contribute to the decay of storms.

The relationships between DV24 and  $S$  are similar for individual ocean basins and for different latitude bands, although the regression slopes for the linear fits between the composite mean DV24 and  $S$  differ with the regions of warmer climatological SST having slightly steeper slopes (Figures S1 and S2). However, the physical mechanism for this sensitivity is not known.

The positive relation between DV24 and  $S$  for intensifying storms suggests that the surplus precipitation could be a good predictor for RI. Our definition of the surplus is analogous to the normalized BT proposed by Fischer et al. (2018) but accentuates its physical meaning, that is, excessive convective heating above that of neutral TCs is approximately proportional to subsequent TC intensity change in terms of composite means averaged over a large number of samples. For individual TCs, many other factors are at play and substantial deviations from the mean relationship are observed, as shown by the joint occurrence frequencies of DV24 and  $S$  (Figure 2b). The correlation between DV24 and  $S$  is 0.24 (statistically significant at 99% level for a sample size over 26,000).

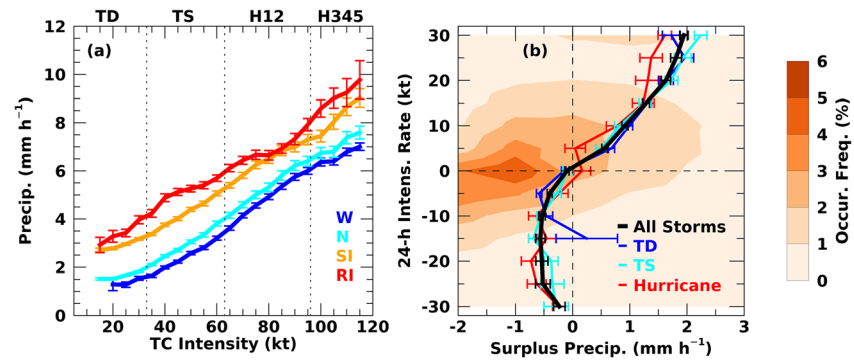
Similar to precipitation, the composite mean IWC from CloudSat and IWP from MERRA-2 exhibit positive correlations with TC intensity and intensification rate (Figures S3 and S4), consistent with the findings in



**Figure 1.** Composite maps of surface precipitation rate (in  $\text{mm hr}^{-1}$ ) in storm-centered coordinate for tropical cyclones in four intensity and four intensification rate groups (see text for details). The precipitation is from TRMM 3B42 from 1998 to 2014.

Wu and Soden (2017). The average IWP within 100 km of storm center is plotted as a function of TC intensity for four intensification groups in Figure S7a. It resembles inner-core precipitation remarkably: It increases roughly linearly with TC intensity for each intensification rate group and the intensifying (weakening) storms have surplus (deficit) IWP over that of neutral TCs of the same intensity. The reason for the striking similarity between inner-core precipitation and IWP is because both quantities are correlated to the magnitude of convective heating (Nolan et al., 2019) and stronger convective heating promote greater TC intensification (Nolan et al., 2007).

Based on TC potential intensity theory, warmer SST and colder outflow temperature would permit stronger maximum intensity and thus higher likelihood of intensification (Emanuel et al., 2013; Wang et al., 2014). Therefore, it is not surprising that TC intensity and intensification rate are positively correlated with SST minus 100-hPa temperature and negatively correlated with 100-hPa temperature from MLS and MERRA-2 (Figures S5–S7). However, the composite TTL temperature differences between the weakening



**Figure 2.** (a) Composite inner-core precipitation rate as a function of TC intensity (in kt) for four intensification groups. (b) TC future 24-hr intensity change (DV24) as a function of surplus precipitation for all TCs and three TC intensity groups. The standard errors of precipitation and the surplus precipitation for each group are shown in (a) and (b), respectively. The joint occurrence frequency of DV24 and the surplus precipitation is shown in orange color shading.

and neutral storms are rather small, indicating that the outflow temperature is not critical in determining the weakening of a TC. Unlike the inner-core precipitation and IWP, the TTL temperature is not a simple linear function of TC current intensity. It increases from tropical depression to tropical storm and then decreases with increasing TC intensity for hurricanes (Figure S7b). The deficit in the TTL temperature relative to that of neutral TCs could be a potential predictor for TC intensification rate.

### 3.2. Using ML Techniques to Apply the Satellite Observations Into RI Forecast

To test the predictive power of the surplus precipitation and other measurements of storm structures, we apply ML techniques for RI forecast with the SHIPS RII predictors and the parameters representing internal storm structures. To compare the ML forecasts with NHC's operational RI consensus, only cases from the Atlantic and EPAC are included in the ML models. The ML models are trained for the Atlantic and EPAC basins separately using the SHIPS developmental database, which are based on Climate Forecast System Reanalysis (CFSR) and GOES IR data from 1998 to 2008. Then we test the model in RI forecasting using the archived National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) forecast fields, GOES BTs, and NHC operational initial storm intensity and location data for the period of 2009 to 2014. Note that the training data uses a “perfect prog” approach where the poststorm best track positions and model analysis fields are used to determine the ML parameters, but the testing data set uses only what was available in real time, including forecast tracks and global model forecast fields. The period of 2009–2014 was chosen because the available NCEP forecast fields start from 2009 and the TRMM 3B42 precipitation data ended after 2014. We compare our ML model results against the NHC operational RI consensus from forecast reruns. The operational RI consensus is the average of the logistic regression, Bayesian, and the discriminant analysis-based SHIPS-RII probabilistic forecast results and is the most skillful model for RI at the NHC according to K15 for 2008–2013 TCs. Three RI thresholds are considered, that is, DV24  $\geq$  25, 30 and 35 kt. Both deterministic and probability forecast skill are assessed.

The predictive skill for deterministic RI forecast are evaluated in terms of three metrics: probability of detection (POD), false alarm ratio (FAR), and the Peirce Skill Score (PSS) (K10; K15). Denoting the number of true positive, false positive, false negative, true negative as  $a$ ,  $b$ ,  $c$ , and  $d$ , respectively,  $POD = \frac{a}{a+c}$ ,  $FAR = \frac{b}{a+b}$ , and  $PSS = \frac{ad-bc}{(a+c)(b+d)}$ , with the best scores being 1, 0, and 1, respectively.

For the probability forecasts, the Brier Skill Score (BSS) relative to the climatology (e.g., K10; K15) is a metric routinely used at the NHC. The Brier score for a forecast model is defined as  $BSM = \frac{1}{N} \sum_{i=1}^N (f_i - O_i)^2$ , where  $f_i$  is the forecast probability and  $O_i$  is 1 or 0 if the event is observed or not, respectively. The BSS relative to the climatology is defined as  $BSS = \left(1 - \frac{BSM}{BSC}\right)$ , whereas the Brier score of the climatological forecast is BSC. A positive value of BSS indicates a better predictive skill than climatological forecast and a BSS of 1



corresponds to a perfect model. In this study, the climatological probability of RI in each basin is based on the values listed in the SHIPS RII forecast dataset for 2009–2014 TCs.

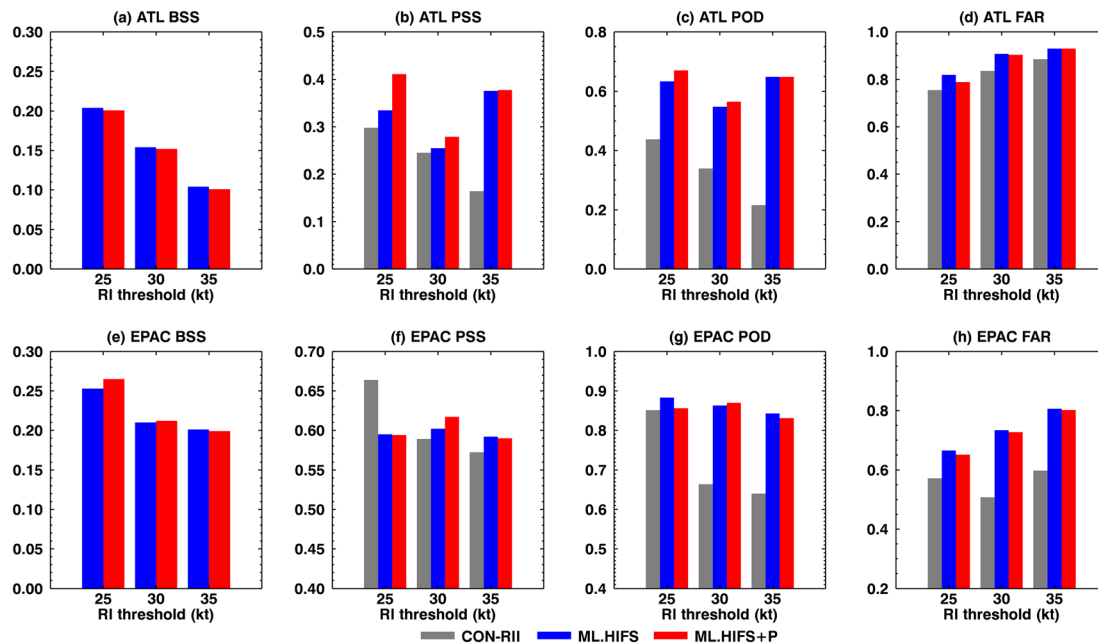
We employ a number of ML schemes, including logistic regression, random forest, decision tree and extra tree available in standard Python scikit-learn package and IBM Watson Studio to build our ML Hurricane Intensity Forecast Scheme (ML.HIFS). A weighted average of the RI probabilities from the four ML schemes is constructed for the final model, which provides a dichotomous RI classifier and the corresponding RI probability (see supporting information). In the training of the ML models using the 1998–2008 samples, we adopt the leave-one-year-out method (K15) to cross validate the model performance, that is, any 10 years of data are used to train a model and the year left out is used to verify the model performance. We compute the BSS and PSS for the training process using all the years that were verified. Then we vary the hyperparameters in individual ML schemes and the weights of different ML schemes in the ensemble models in order to seek a combination of parameters that yield the highest BSS or PSS for the training data. The final ML.HIFS corresponds to the one with the highest BSS in a particular basin and for a specific RI threshold. If we select the final ML.HIFS that yields the highest PSS for the training data, the exact forecast performance for the test data of 2009–2014 could vary but the conclusions are qualitatively similar. When the final ML.HIFS is applied to the test data from 2009 to 2014, we use NCEP forecast fields in place of reanalysis fields. Note that we have only explored a limited number of tunable hyperparameters in the ML schemes as the purpose of this study is to demonstrate a prototype and potential impacts of using ML to incorporate satellite observations into operational forecasts. For real-time operations, the hyperparameters and weights for the individual ML schemes should be updated based on available samples prior to each hurricane season.

Nine predictors employed in the current SHIPS-RII model are shown in Table S4. The details of these predictors are described in K15. We construct ML models using the SHIPS RII predictors with and without the surplus inner-core precipitation from TRMM. The inner-core IWP and 100 hPa temperature from MERRA-2 are also used separately in the ML modeling. CloudSat and MLS measurements are not used in forecast modeling due to limited sample size.

Figures 3a and 3e display the BSS of the final ML.HIFS performance using the SHIPS RII predictors without (blue) and with (red) the surplus precipitation from TRMM based on the leave-one-year-out cross validation for 1998–2008 training data. The corresponding PSS values are shown in Figure S8. The BSS values of our model for the Atlantic (EPAC) are about 20% (25%), 15% (21%), and 10% (20%) for 25-, 30- and 35-kt RI thresholds, respectively, very close to the operational RI consensus model rerun for 2004–13 (see Figure 5 in K15). The BSS values for EPAC are always higher than the Atlantic counterparts, as known for the operational models at the NHC (K10; K15). Using the SHIPS RII predictors only, our ML.HIFS produces PSS for the Atlantic (EPAC) about 0.58 (0.625), 0.605 (0.64), and 0.60 (0.68) for 25-, 30- and 35-kt RI thresholds, respectively. When the surplus inner-core precipitation is combined with the SHIPS RII predictors in the ML.HIFS, the BSSs do not change significantly (Figures 3a and 3e) and the PSSs increase or decrease by 0.1–0.2 in absolute value and 1–3% relatively (Figure S8), probably because the infrared BT predictor in the SHIPS RII already represents some of the inner-core convective structures.

When the ML.HIFS is applied to the test data from 2009–2014, the resulting BSS, PSS, POD, and FAR for RI episodes in the Atlantic and EPAC are compared with the counterparts from the operational RI consensus forecast (Figure 3 and Table S5).

For the Atlantic basin, the PSS, POD, and FAR for 25-kt RI threshold derived from the operational RI consensus forecast probability for 2009–2014 TCs are similar to those in K15 for 2008–2013 TCs, but the PSSs and PODs for 30- and 35-kt RI thresholds are noticeably higher than their counterparts for 2008–2013 TCs in K15, likely because the operational RI consensus model used more data in the regression fitting than that in K15. Using the SHIPS RII predictors only, our ML.HIFS outperforms the NHC consensus forecast for all RI thresholds in the Atlantic basin (Figures 3b and 3c and Table S5). The absolute increases in PSS for 25-, 30-, and 35-kt RI thresholds are 0.04, 0.01, and 0.22, corresponding to relative improvements of 13%, 4%, and 138%, respectively. Such increases in PSS are associated with drastic increases in POD offset by slightly higher FAR. For DV24  $\geq$  35 kt, the ML.HIFS has a POD about 3 times the NHC operational RI consensus and only  $\sim$ 6% higher FAR. With the addition of the TRMM precipitation as a predictor, the PSS for 25-kt RI is further increased to 0.41, a 37% relative improvement. This is associated an increase in POD and a reduction in FAR compared to using the SHIPS RII predictors only. The inclusion of the surplus



**Figure 3.** Predictive skill for RI occurrence in North Atlantic (top) and Eastern North Pacific (bottom) for three RI thresholds (DV24  $\geq$  25, 30, and 35 kt). (a and e) The Brier Skill Score (BSS) relative to climatology for the training data from 1998 to 2008 using the leave-one-year-out cross-validation method. (b–h) The forecast scores for the test data from 2009 to 2014. (b and f) the Peirce Skill Score (PSS), (c and g) Probability of Detection (POD), and (d and h) False Alarm Ratio (FAR). The gray bars represent the operational RI consensus forecast scores and the blue (red) bars represent the machine learning model using the SHIPS RI predictors without (with) the surplus precipitation from TRMM. Better skill corresponds to higher BSS, PSS and POD, and lower FAR.

precipitation in the ML model produces a 12% improvement relative to the NHC model for 30-kt RI. Hence, the ingestion of the surplus precipitation in the ML could further enhance the improvement by the ML relative to the NHC model by 24% (8%) for 25- (30-)kt RI thresholds, but no impact for 35-kt RI threshold. In summary, our ML.HIFS with the surplus precipitation exceeds the operational RI consensus by 37%, 12%, and 138% in PSS; 40%, 60%, and 200% in POD; and 4%, 7%, and 6% in FAR for DV24  $\geq$  25, 30, and 35 kt, respectively. We note that the cost function in our ML model training oversamples the RI cases to maximize the overall predictive skill for RI (see supporting information), resulting in higher POD and FAR than the operational RI consensus. Without the oversampling, the ML model tends to produce lower FAR but also undesirably lower POD and PSS.

For EPAC, the operational RI consensus for 2009–2014 exhibits a substantial improvement from the 2008–2013 scores shown in K15: the PSS values for 25-, 30-, and 35-kt RI thresholds are 0.66, 0.59, and 0.57 respectively, compared to 0.45, 0.50, and 0.42 for 2008–2013 TCs (K15). Our ML.HIFS produces a PSS of 0.60 for 25-kt RI threshold, 9% lower than the operational RI consensus forecast. For 30- and 35-kt RI thresholds, the ML.HIFS produces 2–5% better PSS than the operational RI consensus, associated with simultaneous increases in POD and FAR (Figure 3 and Table S5). Nevertheless, applying ML techniques and including precipitation data could improve the RI forecast relative to the NHC operational model in EPAC, but the relative improvement tends to be small, possibly because the NHC model for EPAC is fairly good and it is more difficult to generate sizeable improvements than in the Atlantic. Adding precipitation would improve the PSS by about 3% for the 30-kt RI and has nearly no impact on the PSS for 25- and 35-kt RI in EPAC.

Previous studies showed that RI under moderate wind shear of  $5\text{--}10\text{ m s}^{-1}$  can be particularly difficult to forecast (Bhatia & Nolan, 2013). We find that our ML model improves the Atlantic RI forecast for the moderate shear cases by about 50%, 15%, and 80% in PSS for 25-, 30-, and 35-kt RI thresholds relative to the operational RI consensus, but very little in EPAC (Figure S9).

Applying the surplus IWP and deficit 100-hPa temperature relative to the neutral TCs as additional predictors in the ML.HIFS, we find varying degrees of improvement or degradation relative to the operational RI

consensus (Figure S10). The degradation of the ML model performance when a new predictor is added is not surprising because overfitting with increased degree of freedom during training may result in poorer performance in testing. It is not clear what would be the optimal degree of freedom for the ML model to produce the best performance. The forecast impacts of inner-core IWP and 100-hPa temperature from MERRA-2 are generally weaker than that of precipitation, partly due to the uncertainties associated with the reanalysis product.

#### 4. Conclusions and Discussions

This study shows that satellite observations of internal storms structures such as inner-core precipitation, IWC/IWP, and outflow temperature bear simple relations with TC intensification rate when sorted by TC current intensity. Using the properties of neutral TCs as thresholds, the surplus or deficit in inner-core precipitation, IWP and TTL temperature are found approximately linearly correlated with subsequent 24-hr TC intensity change. These statistically robust relationships are physically meaningful and form the basis for these satellite measurements to act as predictors for TC future intensity change, especially for RI.

Aiming for rapid application of new satellite observations into operational forecasts, we present a ML framework that combines National Aeronautics and Space Administration (NASA) satellite products and conventional predictors employed at the NHC for statistical RI forecasts. An ensemble average of linear and nonlinear ML models is constructed. We show that the ensemble ML model produces noteworthy improvements for RI forecast in the Atlantic basin, especially for RI of DV24  $\geq 35$  kt. For such intense RI episodes, the NHC model performance has been disappointingly low in the Atlantic (PSS < 0.2) (K10; K15). Our ML.HIFS increases the POD of the 35-kt RI events by  $\sim 200\%$  at the cost of only 6% higher FAR. The overall skill for these RI episodes is increased by 138%. For RI of DV24  $\geq 25$  kt, our ML.HIFS outperforms the NHC best model by 37% and the contribution by adding the surplus precipitation is 24%.

For EPAC, the ML model surpasses the operational RI consensus by a small margin (up to 5%) except for 25-kt RI, whereas the operational RI consensus performed exceptionally well during 2009–2014. The ingestion of MERRA-2 inner-core IWP and the tropopause temperature produces mixed and weaker impacts on RI forecast compared to precipitation. Further study is needed to understand the roles of condensates and outflow temperature.

Our results suggest that intelligently applying satellite retrievals into operational forecast with ML techniques has tremendous potential. The Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) provides precipitation measurements at 0.5 hr temporal and  $0.1^\circ$  spatial resolution over most of the Earth's surface with 3-hr latency, adequate for operational TC forecast guidance (Huffman et al., 2015). With more and more constellations of miniaturized satellites coming into orbit, frequent sampling of storm internal structures and their environments will be possible. A ML-based statistical modeling framework will be extremely valuable to transition these observations into operational use and significantly improve TC intensity forecasts, especially for RI.

#### Conflict of Interest

The authors declare no competing interests.

#### Data Availability Statement

All data used in this study are publicly available. The Best Track tropical cyclone data are available at <https://www.nhc.noaa.gov/data/#hurdat> and <https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks> websites. The TRMM precipitation can be downloaded from <https://pmm.nasa.gov/data-access/downloads/trmm> website. The MERRA-2 reanalysis products are available at [https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data\\_access/](https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/) website. The CloudSat tropical cyclone overpass data set is available at <https://ade-laide.cira.colostate.edu/tc/> website. The MLS temperature data can be downloaded from <http://mls.jpl.nasa.gov> website. The SHIPS developmental and forecast datasets are available at [http://rammb.cira.colostate.edu/research/tropical\\_cyclones/ships/developmental\\_data.asp](http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/developmental_data.asp) website. Please contact the corresponding author Hui Su at [hui.su@jpl.nasa.gov](mailto:hui.su@jpl.nasa.gov) for any questions.



## Acknowledgments

We thank Kerry Emanuel and Todd Julian for useful discussions, and two anonymous reviewers for constructive comments. H. S., L. W., J. H. J., and P. T. performed the work at Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. R. P. was supported by IBM Global Business Services and IBM Watson Analytics. A. L. was supported by RMDS Lab. This study is supported by Aura MLS and CloudSat flight projects, and JPL Data Science Pilot Project.

## References

- Bhalachandran, S., Haddad, Z. S., Hristova-Veleva, S. M., & Marks, F. D. Jr. (2019). The relative importance of factors influencing tropical cyclone rapid intensity changes. *Geophysical Research Letters*, 46, 2282–2292. <https://doi.org/10.1029/2018GL079997>
- Bhatia, K. T., & Nolan, D. S. (2013). Relating the skill of tropical cyclone intensity forecasts to the synoptic environment. *Weather and Forecasting*, 28(4), 961–980. <https://doi.org/10.1175/WAF-D-12-00110.1>
- Blackwell, W., Allen, G., Galbraith, C., Leslie, R., Osaretin, I., Scarito, M., et al. (2013). *MicroMAS: A first step towards a nanosatellite constellation for global storm observation*, 27th Annual AIAA/USU Conference on Small Satellites (pp. SSC13- XI-1). Logan: UT, Utah State University.
- Cecil, D. J., & Zipser, E. J. (1999). Relationships between tropical cyclone intensity and satellite-based indicators of inner core convection: 85-GHz ice-scattering signature and lightning. *Monthly Weather Review*, 127, 103–123. [https://doi.org/10.1175/1520-0493\(1999\)127%3C0103:RBTCA%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127%3C0103:RBTCA%3E2.0.CO;2)
- Chen, H., & Zhang, D. L. (2013). On the rapid intensification of Hurricane Wilma (2005). Part II: Convective bursts and the upper-level warm core. *Journal of the Atmospheric Sciences*, 70, 146–162. <https://doi.org/10.1175/JAS-D-12-062.1>
- Chen, X., Wang, Y., Fang, J., & Xue, M. (2018). A numerical study on rapid intensification of Typhoon Vicente (2012) in the South China Sea. Part II: Roles of inner-core processes. *Journal of the Atmospheric Sciences*, 75, 235–255.
- DeMaria, M., Sampson, C. R., Knaff, J. A., & Musgrave, K. D. (2014). Is tropical cyclone intensity guidance improving? *Bulletin of the American Meteorological Society*, 95(3), 387–398. <https://doi.org/10.1175/BAMS-D-12-00240.1>
- Emanuel, K., Solomon, S., Folini, D., Davis, S., & Cagnazzo, C. (2013). Influence of tropical tropopause layer cooling on Atlantic hurricane activity. *Journal of Climate*, 26, 2288–2301. <https://doi.org/10.1175/JCLI-D-12-00242.1>
- Fischer, M. S., Tang, B. H., & Corbosiero, K. L. (2018). Normalized convective characteristics of tropical cyclone rapid intensification events in the North Atlantic and Eastern North Pacific. *Monthly Weather Review*, 146, 1133–1155. <https://doi.org/10.1175/MWR-D-17-0239.1>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, version-2 (MERRA-2). *Journal of Climate*, 30, 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- Hazelon, A. T., Hart, R. E., & Rogers, R. F. (2017). Analyzing simulated convective bursts in two Atlantic hurricanes. Part II: Intensity change due to bursts. *Monthly Weather Review*, 145, 3095–3117.
- Hendricks, E. A., Peng, M. S., Fu, B., & Li, T. (2010). Quantifying environmental control on tropical cyclone intensity change. *Monthly Weather Review*, 138, 3243–3271. <https://doi.org/10.1175/2010MWR3185.1>
- Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, 8, 38–55. <https://doi.org/10.1175/JHM560.1>
- Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., & Xie, P. (2015). NASA global precipitation measurement Integrated Multisatellite Retrievals for GPM (IMERG). *Algorithm Theoretical Basis Doc., version 4.5*, 4, 26.
- Jiang, H. (2012). The relationship between tropical cyclone intensity change and the strength of inner-core convection. *Monthly Weather Review*, 140, 1164–1176. <https://doi.org/10.1175/MWR-D-11-00134.1>
- Jiang, H., & Ramirez, E. M. (2013). Necessary conditions for tropical cyclone rapid intensification as derived from 11 years of TRMM data. *Journal of Climate*, 26, 6459–6470. <https://doi.org/10.1175/JCLI-D-12-00432.1>
- Jiang, H., Tao, C., & Pei, Y. (2019). Estimation of tropical cyclone intensity in the North Atlantic and northeastern Pacific basins using TRMM satellite passive microwave observations. *Journal of Applied Meteorology and Climatology*, 58, 185–197. <https://doi.org/10.1175/JAMC-D-18-0094.1>
- Kaplan, J., & DeMaria, M. (2003). Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Weather and forecasting*, 18, 1093–1108. [https://doi.org/10.1175/1520-0434\(2003\)018%3C1093:LCORIT%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018%3C1093:LCORIT%3E2.0.CO;2)
- Kaplan, J., DeMaria, M., & Knaff, J. A. (2010). A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Weather and Forecasting*, 25, 220–241. <https://doi.org/10.1175/2009WAF2222280.1>
- Kaplan, J., Rozoff, C. M., DeMaria, M., Sampson, C. R., Kossin, J. P., Velden, C. S., et al. (2015). Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Weather and Forecasting*, 30, 1374–1396. <https://doi.org/10.1175/WAF-D-15-0032.1>
- Kieper, M. E., & Jiang, H. (2012). Predicting tropical cyclone rapid intensification using the 37 GHz ring pattern identified from passive microwave measurements. *Geophysical Research Letters*, 39, L13804. <https://doi.org/10.1029/2012GL052115>
- Molinari, J., & Vollaro, D. (2010). Rapid intensification of a sheared tropical storm. *Monthly Weather Review*, 138, 3869–3885. <https://doi.org/10.1175/2010MWR3378.1>
- Nolan, D. S., & Grasso, L. D. (2003). Nonhydrostatic, three-dimensional perturbations to balanced, hurricane-like vortices. Part II: Symmetric response and nonlinear simulations. *Journal of the Atmospheric Sciences*, 60, 2717–2745. [https://doi.org/10.1175/1520-0469\(2003\)060<2717:NTPTBH.2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<2717:NTPTBH.2.0.CO;2)
- Nolan, D. S., Miyamoto, Y., Wu, S., & Soden, B. J. (2019). On the correlation between Total condensate and moist heating in tropical cyclones and applications for diagnosing intensity. *Monthly Weather Review*, 147, 3759–3784. <https://doi.org/10.1175/MWR-D-19-0010.1>
- Nolan, D. S., Moon, Y., & Stern, D. P. (2007). Tropical cyclone intensification from asymmetric convection: Energetics and efficiency. *Journal of the Atmospheric Sciences*, 64, 3377–3405. <https://doi.org/10.1175/JAS3988.1>
- Rappaport, E. N., Jiing, J.-G., Landsea, C. W., Murillo, S. T., & Franklin, J. L. (2012). 2012: The joint hurricane test bed: Its first decade of tropical cyclone research-to-operations activities reviewed. *Bulletin of the American Meteorological Society*, 93, 371–380. <https://doi.org/10.1175/BAMS-D-11-00037.1>
- Rogers, R. F., Reasor, P. D., & Zhang, J. (2015). Multiscale structure and evolution of Hurricane Earl (2010) during rapid intensification. *Monthly Weather Review*, 143, 536–562. <https://doi.org/10.1175/MWR-D-14-00175.1>
- Rozoff, C. M., Velden, C. S., Kaplan, J., Kossin, J. P., & Wimmers, A. J. (2015). Improvements in the probabilistic prediction of tropical cyclone rapid intensification with passive microwave observations. *Weather and Forecasting*, 30, 1016–1038. <https://doi.org/10.1175/WAF-D-14-00109.1>
- Schwartz, M. J., Lambert, A., Manney, G. L., Read, W. G., Livesey, N. J., Froidevaux, L., et al. (2008). Validation of the aura microwave limb sounder temperature and geopotential height measurements. *Journal of Geophysical Research*, 113, D15S11. <https://doi.org/10.1029/2007JD008783>
- Shapiro, L. J., & Willoughby, H. E. (1982). The response of balanced hurricanes to local sources of heat and momentum. *Journal of the Atmospheric Sciences*, 39, 378–394. [https://doi.org/10.1175/1520-0469\(1982\)039%3C0378:TROBHT%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039%3C0378:TROBHT%3E2.0.CO;2)

- Shimada, U., Owada, H., Yamaguchi, M., Iriguchi, T., Sawada, M., Aonashi, K., et al. (2018). Further improvements to the Statistical Hurricane Intensity Prediction Scheme using tropical cyclone rainfall and structural features. *Weather and Forecasting*, 33, 1587–1603. <https://doi.org/10.1175/WAF-D-18-0021.1>
- Stevenson, S. N., Corbosiero, K. L., & Molinari, J. (2014). The convective evolution and rapid intensification of Hurricane Earl (2010). *Monthly Weather Review*, 142, 4364–4380. <https://doi.org/10.1175/MWR-D-14-00078.1>
- Tourville, N., Stephens, G., DeMaria, M., & Vane, D. (2015). Remote sensing of tropical cyclones: Observations from *CloudSat* and A-Train profilers. *Bulletin of the American Meteorological Society*, 96, 609–622. <https://doi.org/10.1175/BAMS-D-13-00282.1>
- Wang, S., Camargo, S. J., Sobel, A. H., & Polvani, L. M. (2014). Impact of the tropopause temperature on the intensity of tropical cyclones: An idealized study using a mesoscale model. *Journal of the Atmospheric Sciences*, 71, 4333–4348. <https://doi.org/10.1175/JAS-D-14-0029.1>
- Willoughby, H. E. (1990). Temporal changes of the primary circulation in tropical cyclones. *Journal of the Atmospheric Sciences*, 47(2), 242–264. [https://doi.org/10.1175/1520-0469\(1990\)047%3C0242:TCOTPC%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047%3C0242:TCOTPC%3E2.0.CO;2)
- Wu, L., Su, H., Fovell, R. G., Wang, B., Shen, J. T., Kahn, B. H., et al. (2012). Relationship of environmental relative humidity with North Atlantic tropical cyclone intensity and intensification rate. *Geophysical Research Letters*, 39, L20809. <https://doi.org/10.1029/2012GL053546>
- Wu, S., & Soden, B. J. (2017). Signatures of tropical cyclone intensification in satellite measurements of ice and liquid water content. *Monthly Weather Review*, 145, 4081–4091. <https://doi.org/10.1175/MWR-D-17-0046.1>
- Zagrodnik, J. P., & Jiang, H. (2014). Rainfall, convection, and latent heating distributions in rapidly intensifying tropical cyclones. *Journal of the Atmospheric Sciences*, 71, 2789–2809. <https://doi.org/10.1175/JAS-D-13-0314.1>